



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH



Fundació
Miquel Agustí

Desenvolupament de models de predicció de característiques nutricionals del tomàquet mitjançant espectroscòpia d'infraroig proper

Treball de final de grau

Enginyeria de Sistemes Biològics

Autor: Max Michel Todó

Tutor: Joan Simó Cruanyes

Octubre 2018

Agraïments

Gràcies a tots els integrants de la Fundació Miquel Agustí per tota l'ajuda i suport, especialment a l'Helena i la Sílvia que han estat en tot moment pendents de mi amb gran paciència. També als molts altres alumnes que han compartit temps amb mi els hi agraeixo el suport.

Gràcies David, Geri i Rimón per sempre ser allà quant ho he necessitat, mai heu fallat.

Els meus pares sempre han estat incondicionals amb mi, aquesta vegada no ha estat menys, més aviat, ho han estat més.

Resum

El tomàquet (*Solanum lycopersicum* L.) és un dels vegetals més consumits arreu del món i la seva demanda creix anualment. Degut a aquest fet el tomàquet s'ha convertit en un fruit d'especial interès en recerca de noves tecnologies i procediments que en facin més eficient la seva producció.

L'objectiu d'aquest treball és estudiar la capacitat de l'espectroscòpia NIR en la determinació de paràmetres de qualitat del tomàquet a partir de la comparació entre les dades espectrals i les dels anàlisis de referència (anàlisis químics). S'ha estudiat 4 genotips diferents: Guzman 7, Guzman 8, LC748 i Montgrí. Alhora per a cada genotip els fruits han estat cultivats de tres maneres diferents: empeltats amb peu Beaufort, empeltats amb peu Multifort o sense empeltar. Primerament es va fer una lectura amb el tomàquet recent collit del pes i de la quantitat de sòlids solubles totals ($^{\circ}$ Brix) després de enregistrar-ne el seu espectre amb una sonda de fibra òptica. Posteriorment s'ha procedit a fer els anàlisis químics de referència de les mostres on se'n va analitzar: el percentatge de matèria seca, el pH i l'acidesa valorable. Finalment a partir de les mostres deshidratades, resultat de l'estudi de la matèria seca, s'obté una farina homogènia de tomàquet. Aquesta s'enregistra utilitzant el mòdul rotacional de l'espectòmetre NIR per obtenir-ne el perfil spectral. A partir de les dades de referència s'ha pogut estudiar estadísticament les diferències significatives entre genotips i entre els tractaments emprats. Posteriorment s'ha pogut comparar amb els resultats de les dades espectrals per al desenvolupament de models de predicció utilitzant models de regressió PLS.

Els resultats obtinguts dels anàlisis de referència indiquen que hi ha diferències significatives entre els diferents genotips i també entre els tractaments utilitzats. Cada paràmetre estudiat varia però, en general, el pes és la variable amb més desviació entre lectures i el pH i l'acidesa valorable els que menys. L'enregistrament spectral amb la sonda de fibra òptica són correctes però no relacionables amb les dades de referència. Les lectures amb l'esfera amb rotació sí són relacionables amb les dades de referència però és important tenir present que per a fer aquest enregistrament la mostra ha seguit un procés de transformació llarg i costós.

Els models de predicció obtinguts de la relació entre les dades espectrals i les de referència no són bons, és a dir no han assolit els valors dels paràmetres de control establerts per a determinar-los com a vàlids. L'únic model que ha assolit un valor de predicció acceptable és el de predicció de $^{\circ}$ Brix a partir del registre de farina. El contingut en sòlids solubles és dels paràmetres més importants per a la determinació de la qualitat organolèptica d'aquest fruit però aquesta lectura sol fer-se amb el tomàquet sencer.

Paraules clau: *Solanum lycopersicum*, Near Infrared (NIR), models de regressió de mínims parcials quadàtics (PLS), genotip, variància.

Resumen

El tomate (*Solanum lycopersicum* L.) es uno de los vegetales más consumidos alrededor del mundo i su demanda crece anualmente. Debido a este hecho el tomate se ha convertido en una fruta de especial interés en la búsqueda de nuevas tecnologías i procedimientos que hagan de este una producción más eficiente.

El objetivo de este trabajo es estudiar la capacidad de la espectroscopia NIR en la determinación de parámetros de calidad del tomate a partir de la comparación entre los datos espectrales i los de los análisis de referencia (análisis químicos). Se ha estudiado 4 genotipos distintos: Guzman 7, Guzman 8, LC748 i Montgrí. A su vez para cada genotipo los frutos han estado cultivados de tres modos distintos: con injerto con pie de Beaufort, con pie Multifort o sin injerto. Primeramente, se hace la lectura con el tomate recién cogido del peso i de la cantidad de sólidos solubles totales ($^{\circ}$ Brix) después de registrar su espectro con una sonda de fibra óptica. Posteriormente se hace los análisis químicos de referencia de las muestras donde se analiza: el porcentaje de materia seca, el pH i la acidez valorable. Finalmente, de las muestras deshidratadas, resultado del estudio del porcentaje de materia seca de estas, se obtuvo una harina homogénea de tomate. Esta se registra utilizando el módulo rotacional del espectrómetro NIR. Con los datos de referencia se ha podido estudiar estadísticamente las diferencias significativas entre los genotipos i tipos de injertos utilizados. Así ha sido posible comparar los resultados espectrales para el desarrollo de modelos de predicción utilizando modelos de regresión PLS.

Los resultados obtenidos de los análisis de referencia indican que hay diferencias significativas entre los diferentes genotipos i también entre los tratamientos de injerto utilizados. Cada parámetro estudiado varía, pero en general el peso es la variable con más desviación entre lecturas i el pH i la acidez valorable los que menos. Los registros espectrales con la sonda de fibra óptica son correctos, pero no relacionables con los datos de referencia. Las lecturas con la esfera rotacional sí son relacionables con los datos de referencia, pero es importante tener presente que para hacer el registro la muestra ha pasado por un proceso de transformación largo i costoso.

Los modelos de predicción obtenidos de la relación entre los datos espectrales i los de referencia no son buenos, es decir, los valores para los parámetros de control no son óptimos para definirlos como válidos. El único modelo que tiene valores de predicción aceptables es el de predicción de $^{\circ}$ Brix a partir del registro con harina. El contenido en sólidos solubles es uno de los parámetros con más importancia en la determinación organoléptica de este fruto pero en general esta lectura se realiza con el tomate entero.

Palabras clave: *Solanum lycopersicum*, Near Infrared (NIR), modelos de regresión de mínimos cuadrados (PLS), genotipo, varianza.

Abstract

Tomatoe (*Solanum lycopersicum* L.) is once of the most consumed vegetable around the globe and its demand is increasing every year. Due to this fact tomatoe has become an especial fruit in the research of new technologies and procedures that's makes its production more efficient.

The main goal of this work is to study the capacity of NIR spectroscopy on determining quality parametres from comparing between the spectral data and the reference analyses (chemical analyses). Four different geotipes have been studied: Guzman 7, Guzman 8, LC748 and Montgrí. Also, each genotype has been cultivated in different way: with a Beaufort or Multifort grafting or with the absence of it. First, the weight and totals soluble soluts is registered after taking the spectrum of the fruit with an optical prove of the NIR spectrometer. Then the chemistry analyses are done where analyse: the percentage of dry manner, the pH and the titratable acidity. Finally, from the dehydrated samples, resulting from the dry manner percentage study, a homogenous tomatoes flour. This is registered using the rotational sphere of the NIR spectrometer. With the reference data from the chemical study can be studied the significative differences between the genotypes and grafting treatments. Due to the procedure had been possible to compare the spectral data for the development of prediction models using PLS regression models.

The results obtained from the reference analyses indicate that there are significant differences between genotypes and also between grafting treatment. Each parameter studied varies but, in general, weight is the parameter with more deviation on its readings and pH and titratable acidity the less. The spectral registers with the optical prove are corrects but is not possible to relate them with the reference data. The reading with the rotation sphere are related to the reference data but, it is important to have in mind that for doing its registration the fruits have passed through a large and hard process.

The prediction models from the relation of spectral data with reference data are not good, so they are not close enough to the control parameters to accept them as acceptable. The only model that has optimum values is °Brix parameter from the flour registering. The content of soluble solutes is a very important parameter on the organoleptic determination but it is commonly used the fresh and whole fruit for its measurement.

Key words: *Solanum lycopersicum*, Near Infrared (NIR), partial least square regression (PLS), genotype, variance.

ÍNDEX

Índex de figures	7
Índex de taules	9
1. Introducció	10
1.1. Espectroscòpia d'infraroig	10
1.1.1. Fonaments	10
1.1.2. Infraroig proper (NIR)	12
1.1.3. Espectroscòpia FT-NIR	12
1.2. Aplicacions de l'espectroscòpia NIR	13
1.3. El tomàquet	14
1.3.1. Importància socio-econòmica	14
1.3.2. Qualitats químiques del tomàquet	15
1.3.3. Estudis amb NIR del tomàquet	16
2. Objectius	17
3. Materials i mètodes	18
3.1. Material vegetal	18
3.2. Disseny experimental	18
3.3. Anàlisi de referència	19
3.4. Registre amb el NIR	20
3.5. Anàlisi estadístic	22
4. Resultats i discussió	24
4.1. Dades de referència	24
4.1.1. Estadístiques bàsiques	24
4.2. Anàlisi de l'espectre NIR	30
4.2.1. Espectres sense tractar	30
4.2.2. Regressió de PLS	32

5. Conclusions	34
6. Bibliografia	35

Índex de figures

Figura 1	Propagació d'ona electromagnètica. L'eix x i y representen el camp elèctric i el magnètic respectivament.....	10
Figura 2	Energia potencial en el model d'oscil·lador harmònic.....	11
Figura 3	Tipus de vibracions fonamentals d'una molècula triatòmica.....	11
Figura 4	Bandes d'absorció en la regió NIR.....	12
Figura 5	Esquema d'un aparell d'espectroscòpia FT-IR.....	13
Figura 6	Representació gràfica de l'evolució de la producció mundial del tomàquet des del 1980 fins al 2016.....	14
Figura 7	Producció de tomàquet per països l'any 2015.....	15
Figura 8	Plantació de tomàquet en un dels hivernacles de Vilassar de Mar.....	18
Figura 9	Diagrama dels diferents genotips i tractaments a estudiar. El nombre de repeticions o fruits està indicat entre parèntesis fent referència al codi de color per a cada genotip.....	18
Figura 10	Refractòmetre Atago.	19
Figura 11	pH-metre Crinson utilitzat en la mesura pH i càlcul de l'acidesa valorable.	20
Figura 12	Antaris II FT-NIR Analyzer.	21
Figura 13	Sonda de fibra òptica integrada a l'espectròmetre d'infraroig proper FT-NIR Analyzer.	21
Figura 14	Farina de tomàquet.....	22
Figura 15	Histograma i corba de distribució normal dels °Brix, el percentatge en matèria seca i el pH.....	24
Figura 16	Distribució de les mitjanes i el rang des valors de percentatge en matèria seca en relació al genotip i tractament.....	25
Figura 17	Distribució de les mitjanes i rang de valors de les mesures de pH en relació al genotip i tractament.....	26
Figura 18	Distribució de les mitjanes i rang de valors de les mesures d'acidesa avalorable (TA) en relació al genotip i tractament.....	27
Figura 19	Distribució de les mitjanes i rang de valors de les mesures de pes en relació al genotip i tractament.....	28
Figura 20	Distribució de les mitjanes i rang de valors de les mesures de sòlids solubles (°Brix) en relació al genotip i tractament.....	29

Figura 21	Espectres sense tractar de les mostres molturades diferenciades per genotip i tractament.....	31
Figura 22	Espectres sense tractar de les mostres fresques diferenciades per genotip i tractament.....	31
Figura 23	Representació gràfica de la distribució de mostres segons el procediment de regressió per PLS per la variable °Brix segons genotip i tractament per al registre en fresc o en farina.....	32

Índex de taules

Taula 1	Valor orientatiu de la composició de tomàquet madur.	16
Taula 2	Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures de matèria seca en relació al genotip i tractament.	25
Taula 3	Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures de pH en relació al genotip i tractament.	26
Taula 4	Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures d'acidesa valorable (TA) en relació al genotip i tractament.	27
Taula 5	Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures de pes en relació al genotip i tractament.	28
Taula 6	Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures de sòlids solubles totals (°Brix) en relació al genotip i tractament.	29
Taula 7	Síntesi dels resultats obtinguts pel procediment ANOVA per a l'estudi entre tractaments d'empeltat. Cada lletra ("a", "b" i "c") són els grups de significació, és a dir grups amb diferències delimitadores per al procediment, de manera que es pot establir ràpidament la capacitat de definició dels paràmetres respecte de les variables. La última fila presenta els valors de significació (valor p) dels factors considerats calculant la mitjana d'error quadràtic (<i>Mean Square Error</i> , la mitjana de les distàncies quadràtiques de cada mesura a la mitjana del grup).	30
Taula 8	Síntesi dels resultats obtinguts pel procediment ANOVA per a l'estudi entre tractaments d'empeltat. Cada lletra ("a", "b" i "c") són els grups de significació, és a dir grups amb diferències delimitadores per al procediment, de manera que es pot establir ràpidament la capacitat de definició dels paràmetres respecte de les variables. La última fila presenta els valors de significació (valor p) dels factors considerats calculant la mitjana d'error quadràtic (<i>Mean Square Error</i> , la mitjana de les distàncies quadràtiques de cada mesura a la mitjana del grup).	30
Taula 9	Síntesi dels resultats obtinguts pel procediment de regressió de mínims quadràtics parcials (PLS regression).	33

1. Introducció

1.1. Espectroscòpia d'infraroig

1.1.1. Fonaments

La radiació d'infraroig és la part de l'espectre electromagnètic amb longitud d'ona entre 780 nm i 10000 nm ($13000-1000\text{ cm}^{-1}$). S'anomena amb el terme infraroig, per sota del vermell, ja que el vermell és la part de l'espectre de llum visible amb major longitud d'ona. El descobriment de l'espectre del infraroig s'atribueix a Sir William Herschel l'any 1800 mentre estudiava les variacions de la temperatura de la difracció solar utilitzant un prisma. En aquest estudi va poder observar com a la zona contigua del feix de llum vermella, on no s'apreciava radiació visible, l'augment de temperatura era considerable. Les primeres aplicacions van arribar molt més tard quant els mètodes espectroscòpics començaven a utilitzar-se per a la resolució de problemes analítics (Abney & Festing, 1881). Més endavant van començar a enregistrar-se espectres resultants de la interacció de la radiació d'infraroig amb la matèria orgànica (Norris & Butler, 1961).

El conjunt de radiacions electromagnètiques es coneix com a espectre electromagnètic. En l'espectre de la radiació d'infraroig s'hi poden diferenciar tres regions principals: l'infraroig llunyà (*far infrared*, FIR) entre 10000-4000 nm, el mitjà entre 4000-2500 nm (*mid infrared*, MIR) i el proper, entre 2500-750 nm (*near infrared*, NIR) (Ciurczak, 2007).

Així com la llum visible, l'infraroig és un tipus de radiació electromagnètica que es propaga seguint les propietats físiques de les ones. Aquesta és energia que es propaga a través del buit o a través d'un medi amb el que interaccionarà el seu camp magnètic. D'aquesta manera és possible estudiar les propietats de les ones electromagnètiques per la seva interacció amb la matèria. Les ones electromagnètiques són transversals i els vectors elèctric i magnètic són perpendiculars a la seva direcció de propagació, mutualment perpendiculars i en fase (Figura 1).

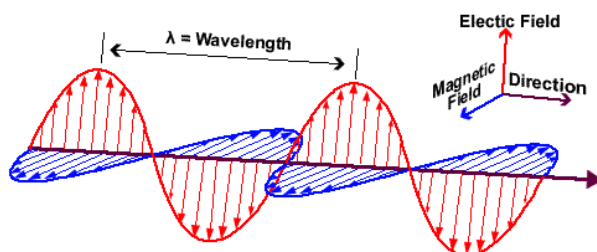


Figura 1. Propagació d'ona electromagnètica (Lin, 2015).

El àtoms d'una molècula mai estan quietes i les vibracions d'aquest àtoms són en direccions específiques depenent de la conformació dels seus enllaços. Hi ha diferents modes vibracionals proporcionals a l'energia potencial de l'enllaç. Quan una ona electromagnètica de l'infraroig interactua amb la matèria es produeixen transicions entre modes vibracionals de diferents nivells energètics, és a dir que incrementa o disminueix la freqüència de vibració de la molècula. El model d'oscil·lador harmònic per a una molècula diatòmica és el més senzill per a representar aquest fenomen físic de manera que el moviment i longitud de separació dels dos àtoms variarà depenent del seu nivell energètic (Figura 2).

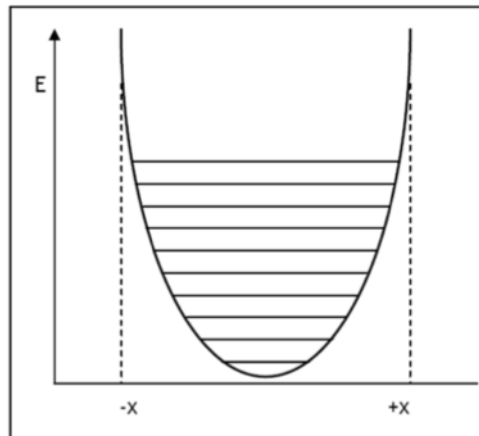


Figura 2. Energia potencial en el model d'oscil·lador harmònic (Peguero, 2010).

Però, aquest model no s'ajusta al comportament real de les molècules i només és representable per aquelles que són diatòmiques. La mecànica clàssica no pot descriure de forma completa el comportament dels àtoms i és necessari tenir en compte alguns conceptes de la física quàntica, on s'utilitza el concepte de nombre quàntic vibracional. Aquests són conjunts vibracionals discrets (ν) amb transicions consecutives definides per Erwin Schrödinger i són específics per a cada molècula degut a, com va demostrar Max Planck, que cada ona electromagnètica té una energia associada per a la seva pertinent longitud d'ona (Skoog, 2001).

En les molècules poliatòmiques les vibracions no només afecten a la longitud sinó també a l'angle de l'enllaç entre àtoms. Diferenciem entre dos tipus de vibracions principals: les vibracions d'extensió (ν) que es troben al llarg de l'enllaç i les vibracions que canvien l'aspecte de la molècula, anomenades vibracions de deformació (δ). A la figura 3 pot observar-se els diferents tipus de moviments que pot ocasionar les vibracions moleculars fonamentals.

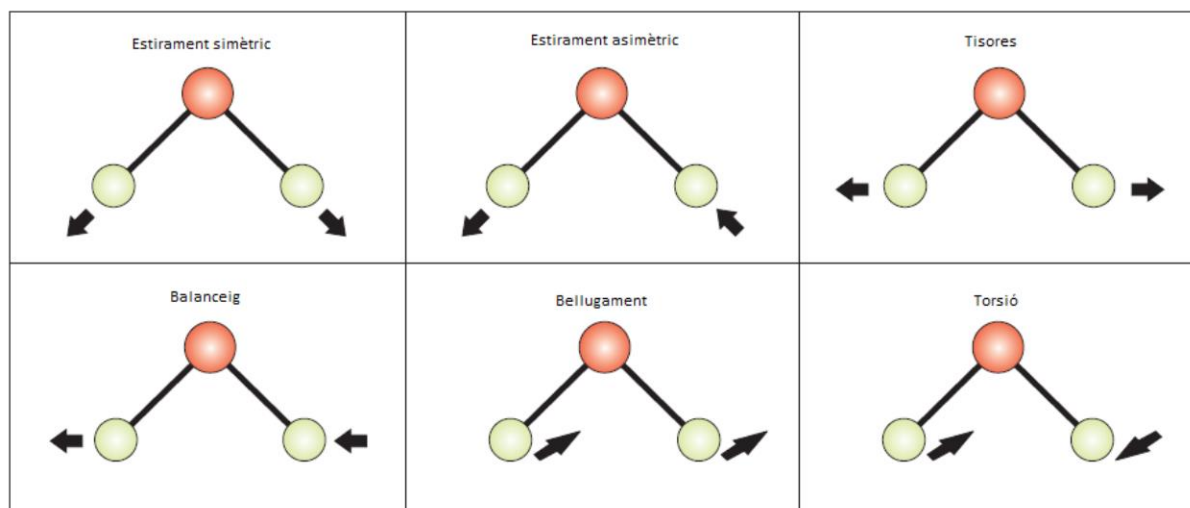


Figura 3. Tipus de vibracions fonamentals d'una molècula triatòmica (Plans, 2014).

D'aquesta manera els salts en l'estat vibracional on es compleix $\Delta\nu = \pm 1$ s'anomenen transicions fonamentals, els salts corresponents $\Delta\nu = \pm 2, \pm 3$ s'anomenen sobretons. Cal tenir en compte que els salts $= \pm 3$ són molt rarament observats. Degut a l'energia associada a cada longitud d'ona s'observa que en el NIR, que és on es poden observar els sobretons i bandes de combinacions de transicions fonamentals, s'ocasionen vibracions moleculars fonamentals (Workman Jr., 1995).

1.1.2. Infraroig proper (NIR)

Tal i com s'ha esmentat, la radiació NIR és la regió de l'espectre electromagnètic que comprèn l'interval entre 750 i 2500nm. Les dues característiques més importants de l'espectre NIR són la presència de sobretons i de bandes de combinació. Els sobretons són aproximadament múltiples de freqüències vibracionals fonamentals i les bandes de combinació són la suma de diferents energies de vibració, siguin fonamentals o sobretons. La probabilitat de les bandes de combinació i sobretons és menor que la de les transicions fonamentals i és per això que la intensitat de l'absorció d'energia per part dels enllaços en la regió NIR és menor que en la regió del MIR, la intensitat d'absorció depèn de l'anharmonicitat de l'enllaç, és a dir la distància d'aproximació del sistema a un d'oscil·lació harmònic. És per això que els enllaços entre àtoms de pesos atòmics molt diferents són els més fàcils de discernir, són més absorbents, amb el NIR ja que la seva anharmonicitat és molt major (Jordan, Edward C., 1968). A la figura 4 es mostren els principals enllaços absorbents al NIR i la seva corresponent regió espectral.

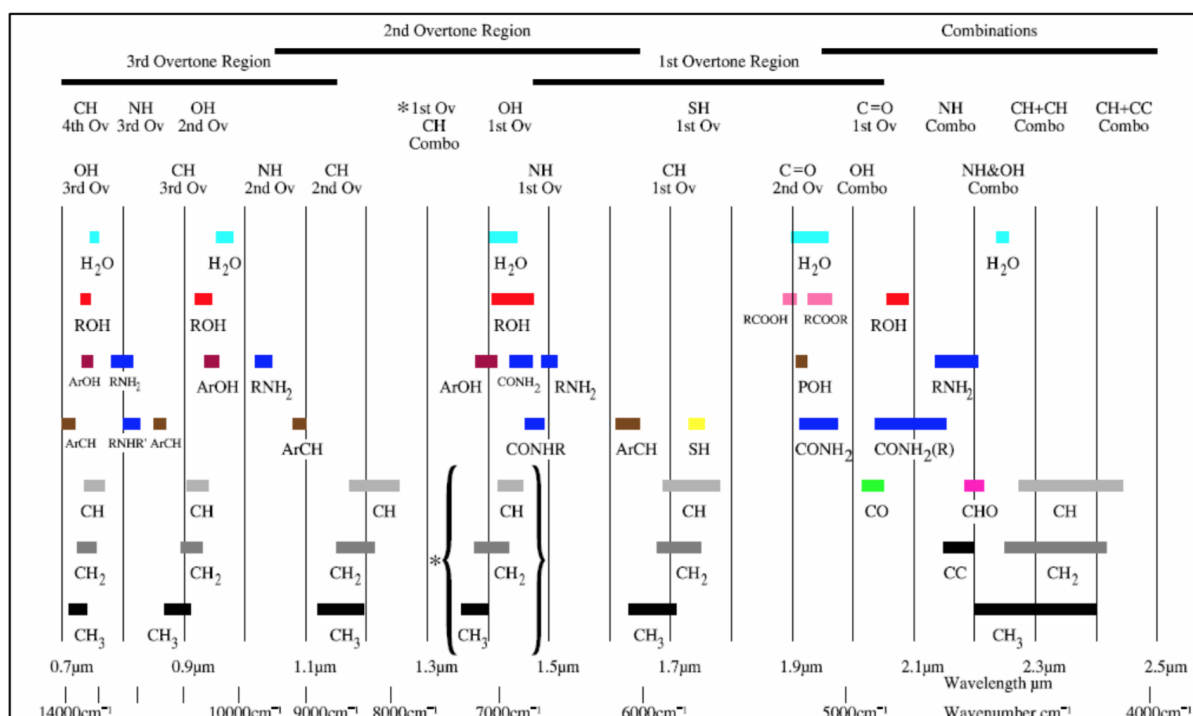


Figura 4. Bandes d'absorció en la regió NIR (Massart, 2003).

1.1.3. Espectroscòpia FT-NIR

L'espectroscòpia d'infraroig proper amb Transformada de Fourier (FT-NIR de l'anglès *Fourier Transform near infrared*) té com a objectiu mesurar de quina manera una mostra absorbeix la llum en diferents espectres concrets utilitzant la Transformada de Fourier com a eina matemàtica per a interpretar els resultats obtinguts. L'espectroscòpia FT-IR requereix de la utilització d'un interferòmetre de Michelson. La interferometria és una família de tècniques on les ones electromagnètiques se superposen causant interferències en la seva radiació de les que se'n poden extreure informació (Stauffer & Belmes, 2018). Gràcies a la implementació d'aquesta tecnologia és possible estudiar un espectre amb moltes freqüències simultànies.

Un equip d'espectroscòpia FT-NIR té fonamenta en una font de llum, el *beamsplitter* (de l'anglès "separador de feixos"), un mirall fix, un de mòbil i els sensors encarregats de fer les mesures òptiques (Baker et al., 2014) (Figura 5).

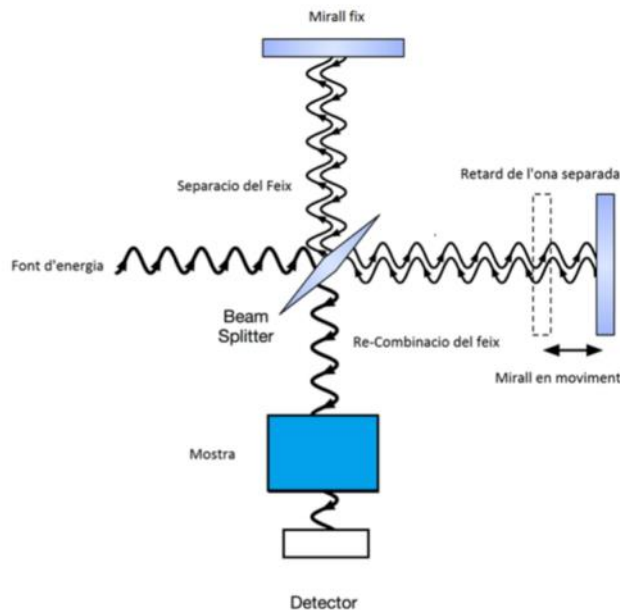


Figura 5. Esquema d'un aparell d'espectroscòpia FT-IR.

La font emet un feix de llum, l'espectre del qual és calibrat amb l'ajut d'un làser, separat en dos al creuar el *beamsplitter* i adreçats cap als miralls. El mirall mòbil és en direcció recta a la font i el fixe a 90 graus del *beamsplitter*. D'aquesta manera s'aconsegueix desfasar la senyal de manera que amb el moviment del mirall sigui possible controlar el nivell de desfasament entre els dos feixos definint així, mitjançant la Transformada de Fourier, la longitud d'ona específica en cada moment.

1.2. Aplicacions de l'espectroscòpia NIR

Les primeres aplicacions de l'espectroscòpia NIR no van ser fins a mitjans del s. XX en que es va estudiar els principals components d'alguns productes orgànics (Kaye, 1954). A partir d'aquest primer tracte amb la tecnologia NIR un equip d'investigació del Departament d'Agricultura dels Estats Units d'Amèrica (*United States Department of Agriculture, USDA*) van esdevenir dels primers impulsors en l'ús d'aquesta tècnica al publicar diferents estudis sobre la composició química i característiques físiques de diferents llavors i grans (Norris, 1964).

Cal tenir en compte que en aquella època no hi havia la possibilitat de fer un registre de manera continua de les mostres degut a la falta de capacitat de processament d'informació i no va ser fins als anys 70, amb l'aparició de nous equips d'espectroscòpia NIR i sobretot de les computadores personals, que aquesta tècnica va guanyar popularitat i va diversificar-se en diferents sectors (Blanco & Villarroya, 2002).

A diferència d'altres regions de l'espectre infraroig, com podria ser el MIR, la recerca amb el NIR ha estat relativament lenta en part perquè molts especialistes la consideraven una regió molt confusa degut al solapament de pics, fent-ne l'assignació molt complexa.

A finals del 2017 el nombre de publicacions on hi ha present el terme espectroscòpia d'infraroig proper és de 28720 segons la base de dades Scopus. Referent a aquesta informació és fàcil entendre que l'augment de la recerca utilitzant aquesta tècnica ha crescut moltíssim però de forma recent ja que el nombre de documents publicats utilitzant-la l'any 2000 era de 420 publicacions però aquest nombre augmenta a més de 2000 per any des del 2012 (Khuwijtjaru, 2018).

En general l'espectroscòpia NIR és una eina amb nombrosos avantatges enfront de pocs inconvenients. Principalment cal destacar que permet el registre de mostres en estat sòlid, líquid o gasos i no és necessari fer un pretractament a aquestes per a fer-ne el registre. A més és una tècnica no destructiva, de rutina d'anàlisi ràpid i econòmic permetent així el registre en línia de producció. A partir d'un sol espectre és possible identificar diferents substàncies tenint present que el registre espectral és relativament ràpid (depèn del tipus d'anàlisi que es vulgui fer). D'altra banda, en contraposició als avantatges, es requereix d'eines matemàtiques molt potents per a interpretar els resultats i sobretot per a modelitzar-los, sovint aquestes eines són programes informàtics de difícil accés per a la majoria. Per altre banda, la inversió inicial és considerable ja que els equips d'espectroscòpia NIR tenen un preu elevat, a més que es requereix de personal especialitzat per a la calibració de l'equip i el seu correcte funcionament. També és necessari tenir el coneixement per a interpretar els resultats obtinguts recordant les peculiars bandes d'absorció en l'espectre NIR i en les bandes solapades i també tenint present quins són els enllaços que ressonen en aquest espectre. Finalment cal tenir en compte que no es poden analitzar aliments que tinguin una concentració d'aigua superior al 40% (Murray, 2004).

1.3. El tomàquet

1.3.1 Importància socio-econòmica

El tomàquet (*Solanum lycopersicum* L.) és el segon vegetal més consumit arreu del món industrialitzat. La seva demanda ha augmentat anualment a nivell mundial de manera constant en què la producció ha augmentat més de 3 vegades des de l'any 1980 (Figura 6)(FAOSTAT,1980-2016). D'aquesta manera la producció mundial va ser al 2014 de més de 170 milions de tones augmentant progressivament fins als 177 milions de tones a l'any 2016 (FAOSTAT, 2014, FAOSTAT, 2016).

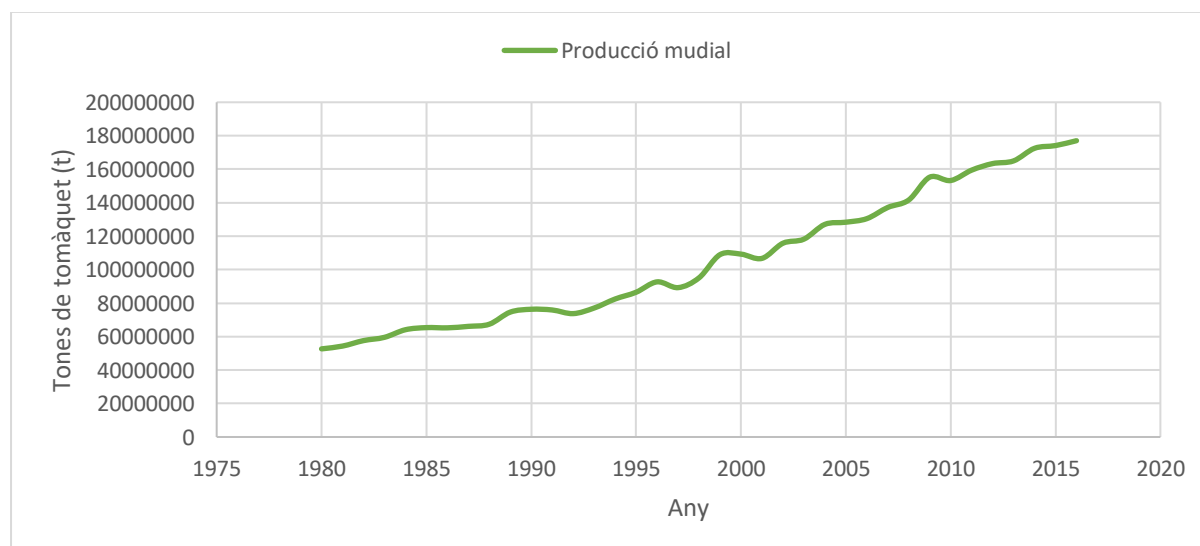


Figura 6. Representació gràfica de l'evolució de la producció mundial del tomàquet des del 1980 fins al 2016 (FAOSTAT,1980-2016).

El primer productor de tomàquet és la Xina amb una producció l'any 2015 de 103 milions de tones aproximadament, representat el 31,5% de la producció mundial. A la figura 6 es pot observar la producció dels 10 països capdavaners en la producció de tomàquet (FAOSTAT, 2015).

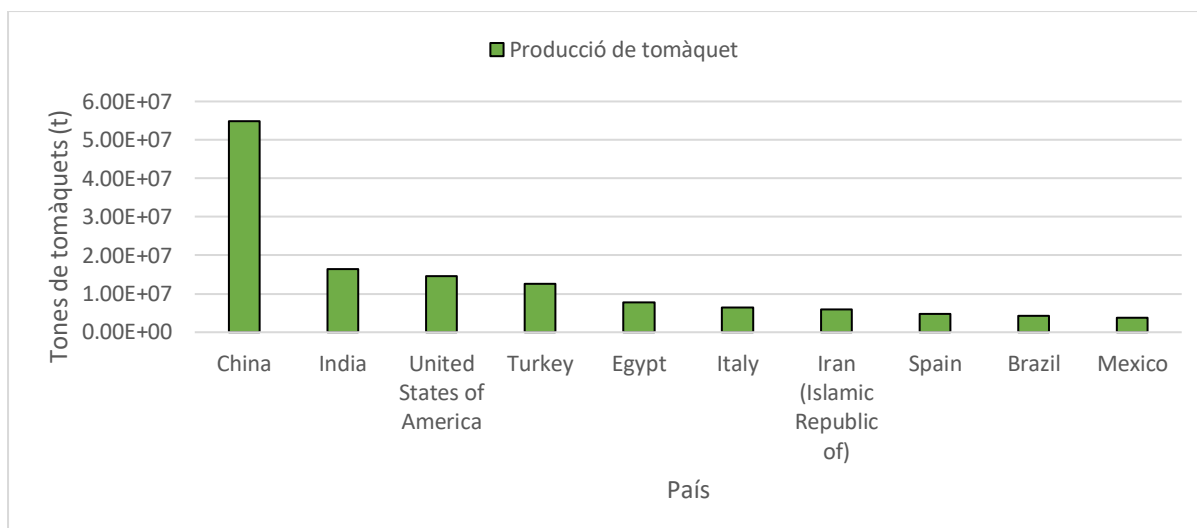


Figura 7. Producció de tomàquet per països l'any 2015 (FAOSTAT,2015).

1.3.2. Qualitats químiques del tomàquet

El tomàquet és una espècie model perquè es consumeix a tot el món i també perquè és un aliment sa. Per això és un dels fruits més estudiats al món. A més és un component clau en la dieta Mediterrània, una de les dietes més amplament reconegudes com a saludables. El tomàquet es caracteritza pel seu alt contingut en aigua, superior al 93%, sucres i sals minerals, especialment potassi (Frusciante et al., 2007). És rellevant el seu contingut en licopè, carotenoide aportador del característic color vermell del tomàquet, acceptat com a beneficiós per les seves propietats com a antioxidant. Addicionalment el seu consum redueix les possibilitats de patir algunes malalties degeneratives així com alguns tipus de càncer (Engelhard et al., 2006). El consum del tomàquet s'ha estès a tot el món on cada cop es té més consciència de la importància d'una dieta sana i equilibrada sent-ne el tomàquet un excel·lent aliment per a qualsevol dieta.

Des del punta de vista nutricional el tomàquet és un producte molt interessant ja que és un aliment ric en carotens, especialment en licopè, a més de vitamines i sals minerals, amb especial elevat contingut en potassi. Per altre banda és un fruit poc energètic que aporta menys de 20Kcal cada 100g degut al seu baix contingut en matèria seca i greixos i és que el 95% del fruit és aigua aproximadament.

El segon major component del tomàquet són els carbohidrats on la glucosa i la fructosa en són els principals amb una proporció d'entre 1,7% i 4% del pes fresc del fruit. Pel que fa a al contingut de minerals els majors representats són el fòsfor, el nitrogen i el potassi amb una concentració del 0,4%, 0,6% i 3% de la matèria seca, respectivament. Les vitamines més abundants són A, B1, B2, B5 i C amb màxima representació de la vitamina A. El valors del pH oscil·la entre 4,0 i 4,8, on l'àcid cítric és el predominant seguit de l'àcid màlic. L'acidesa d'un tomàquet serà una de les propietats més importat del seu perfil sensorial (Nuez, 1995).

Taula 1. Valor orientatiu de la composició de tomàquet madur (Nuez, 1995).

Component	Pes fresc (%)
Matèria seca	6,50
Carbohidrats totals	4,70
Greixos	0,15
Proteïnes	1,00
Fibra	0,50
Sucres reductors	3,00
Sòlids solubles totals (°Brix)	4,50
Àcid màlic	0,10
Àcid cítric	0,20
Potassi	0,25
Vitamina C	0,02

1.3.3. Estudis amb NIR del tomàquet

El sector agroalimentari ha estat dels primers en adoptar l'espectroscòpia NIR com a eina per a fer estudis. Va ser en l'estudi de composició química i percentatge d'humitat en gra i llavors on es van fer els primer passos de recerca amb aquesta eina. Més tard, la possibilitat de conèixer el contingut de proteïna, grassa, fibra, etc, de manera ràpida i econòmica va estimular-ne la seva posta en escena. Manipular informació referent a les propietats químiques i físiques a temps real va ser la característica determinant per a decidir-se d'invertir-hi a gran escala. La capacitat d'obtenir una gran quantitat d'informació, emmagatzemar-la i estudiar-la de manera ràpida, gràcies als avenços tecnològics en el sector computacional i electrònic, van fer d'aquesta, una eina indispensable per al control en producció. A més en els últims anys hi hagut varis escàndols sobre la qualitat d'alguns aliments de consum humà que han posat de manifest la falta d'eines per al control de qualitat en algunes plantes processadores. L'adulteració de carn i greixos pot ser detectat fins al mínim nivell de contaminació d'un 10% (Huck, 2017).

Hi ha molts estudis en el sector agroalimentari sobre el tomàquet que han optat per a la utilització de l'espectroscòpia NIR. S'ha utilitzat amplament per a la determinació de paràmetres de qualitat químics com són el contingut en sòlids solubles, pH, acidesa valorable (Ecarnot et al., 2013) i carotenoides, sent el licopè el principal estudiat (Deák et al., 2015) (Saad et al., 2016). L'espectroscòpia NIR també és útil per a estudiar paràmetres de qualitat físics com són la fermesa (Lee et al., 2014) o el nombre d'esquerdes cuticulars (Huang et al., 2018).

2. Objectius

L'objectiu principal d'aquest estudi és avaluar la capacitat de l'espectroscòpia d'infraroig proper com a eina per a mesurar les característiques de qualitat de tomàquets.

Com a objectius específics, es volen crear models de predicció per als paràmetres químics estudiats: contingut en sòlids solubles, contingut en matèria seca i acidesa valorable a partir del registre per una banda dels fruits sencers i per altre de les mostres en forma de farina (mostra seca i molturada). A més es mira d'establir relacions entre els resultats obtinguts amb l'espectroscòpia NIR i els resultats obtinguts dels anàlisis de referència.

3. Materials i mètodes

3.1. Material vegetal

Els tomàquets utilitzats en aquest estudi van ser cultivats sota túnel a la finca “Del Pagès a Casa” gestionada pel Dalmacio a Vilassar de Mar (41°30'57.5"N2°23'25.9E) (figura 8).



Figura 8. Plantació de tomàquet en un dels hivernacles de Vilassar de Mar. Fotografia: Max Michel

Es van collir tomàquets de 4 genotips diferents cadascun amb 3 tipus de tractament, amb empelt i en absència d'aquest (figura 9). En el tractament d'empeltat es van utilitzar porta-empelts Beaufort i Multifort degut al seu ampli ús entre horticultors professionals.

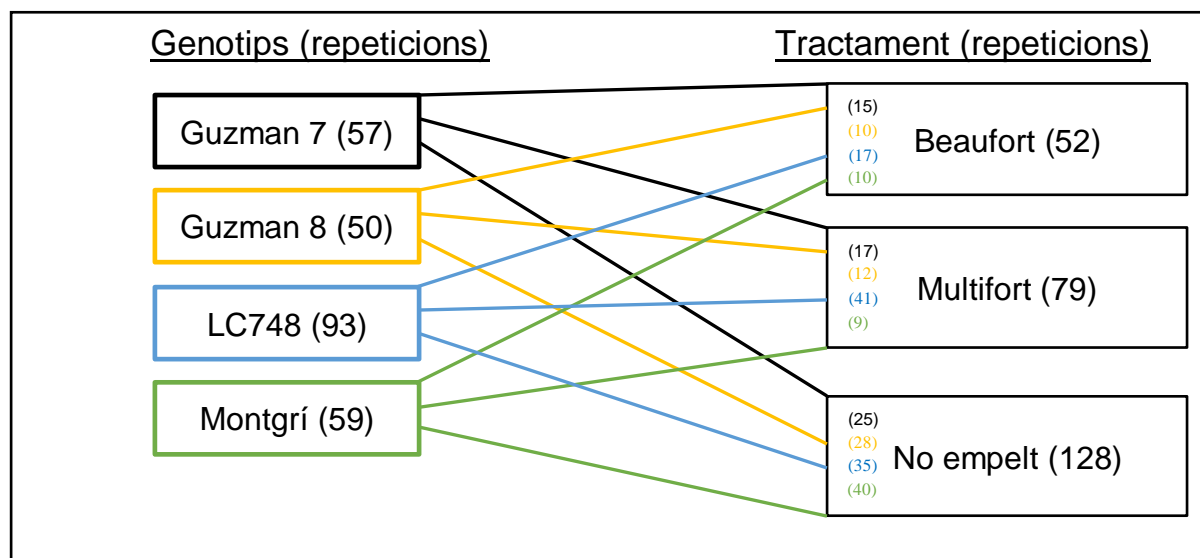


Figura 9. Diagrama dels diferents genotips i tractaments a estudiar. El nombre de repeticions o fruits està indicat entre parèntesis fent referència al codi de color per a cada genotip.

3.2. Disseny experimental

En aquest estudi les mostres, 259 tomàquets (*Solanum lycopersicum* L.) en total, s'han tractat sempre de manera individual. Partint del material vegetal definit en l'apartat “Matèria vegetal”, es mesura el pes, els °Brix i l'espectre dels fruits frescs. Un cop obtingudes les dades referents

a la fase de tractament amb el producte en fresc els fruits s'emmagatzemen congelats per a la posterior manipulació i anàlisi químics.

De les mostres congelades se'n estudia l'acidesa valorable, el pH i el seu contingut en matèria seca. Posteriorment es transformen les mostres deshidratades en farina de manera que se'n estudi, de les mostres en pols, el seu espectre d'infraroig proper.

A partir de la informació recopilada de les mostres en fresc, descongelades i en pols es durà a terme un tractament i anàlisi estadístic per a la síntesi de models matemàtics que puguin relacionar les diferents variables física-químiques amb els espectres obtinguts amb l'espectròmetre d'infraroig proper.

3.3. Anàlisi de referència

Per als anàlisis de referència durant el primer i segon dia d'haver obtingut el material vegetal aquest, en la seva totalitat i després de ser rentat amb aigua destil·lada, va ser pesat utilitzant una balança amb una precisió de 0,01g. A continuació es va tallar en parts cada fruit per a mesurar-ne els °Brix utilitzant un refractòmetre Atago (figura 10) i posar les parts de cada fruit en bosses individuals de plàstic prèviament identificades amb el genotip, tipus de tractament i nombre de mostra per al seu emmagatzematge en cambra a menys -20°C. Al congelar el material es permet de fer-ne posteriors anàlisis tenint una pèrdua mínima de propietats químiques en el temps necessari d'emmagatzematge d'aquest.



Figura 10. Refractòmetre Atago.

Els anàlisis químics comprenen la mesura del percentatge de matèria seca, acidesa valorable (TA) i pH. Primerament les mostres es treuen de la cambra a -20°C i es posen en una cambra a 4°C el dia abans de la seva manipulació per d'aquesta manera descongelar el material lentament sense malmetre'l. A continuació, mitjançant un molinet, es tritura la mostra per produir-ne un puré homogeni. Per a mesurar la matèria seca, es posen 50g aproximadament en una safata d'alumini i es col·loca a l'estufa a 60°C durant 3 dies per evaporar-ne la totalitat del seu contingut en aigua. Per tant, el percentatge de matèria seca és el que resta de la mostra després d'haver-ne eliminat l'aigua. La lectura del percentatge de matèria seca és la mitjana de les dues repeticions fetes per a cada fruit. D'aquesta manera s'estableix la relació entre el pes del fruit en estat fresc i en estat sec mitjançant l'equació 1. Les mostres ens les safates d'alumini sempre es van traslladar d'una sala a una d'altre (de la sala on hi ha l'estufa a la sala de balances) utilitzant un recipient dessecador per a no afectar a les mesures d'humitat de les mostres amb la humitat ambiental.

$$Eq. 1: \%MS = \frac{P_{safata+sec} - P_{cap}}{P_{safata+fresc} - P_{cap}} \cdot 100$$

Per a la mesura del pH i acidesa valorable es disposa 5 grams en un tub Falcon. L'acidesa valorable és la quantitat d'àcids majoritaris (àcid cítric en aquest cas) per cada 100g de mostra. En aquest estudi es va mesurar mitjançant una valoració àcid-base amb NaOH 0,1M, aquest prèviament valorat amb Biftalat de potassi de concentració 0,05M. La mesura de pH i acidesa valorable es fan consecutivament disposant la mostra en un vas de precipitats de 100 ml de volum total i afegint 45 ml d'aigua destil·lada. En constatació mitjançant un agitador magnètic, se'n mesura el pH amb pH-metre Crison (figura 11) i a continuació se'n valora l'acidesa afegint NaOH 0,1M fins arribar a pH 8.1. El volum de NaOH utilitzat s'utilitza per a calcular l'acidesa valorable amb l'equació mostrada a continuació (Eq. 2). Per a cada fruit es van fer dues repeticions de manera que el valor utilitzat per a l'estudi estadístic és el resultat de la mitjana de les dues repeticions.

$$Eq. 2: TA = \frac{V \text{ mL } S_{NaOH}}{m \text{ g mostra}} \cdot \frac{1 \text{ L } S_{NaOH}}{1000 \text{ mL } S_{NaOH}} \cdot \frac{c \text{ mol NaOH}}{1 \text{ L } S_{NaOH}} \cdot \frac{1 \text{ mol cítric}}{n \text{ mol NaOH}} \cdot \frac{M \text{ cítric}}{1 \text{ mol cítric}} = \frac{g \text{ cítric}}{100 \text{ g mostra}}$$



Figura 11. pH-metre Crison utilitzat en la mesura pH i càlcul de l'acidesa valorable. Fotografia: Helena Isern

3.4. Registre amb el NIR

Per a l'estudi de l'espectre electromagnètic del tomàquet es fa dos tipus de registre. Per una banda s'enregistra els tomàquets sencers i per l'altre es mesura les mostres deshidratades molturades. Per fer les lectures de l'espectre electromagnètic s'utilitza un espectròmetre de transformada de Fourier d'infraroig proper, Antaris II FT-NIR Analyzer (figura 12). L'aparell té tres mòduls diferenciats per a possibles necessitats específiques de les diferents mostres a enregistrar. Es fa servir el programari informàtic propi de l'aparell per a dur a terme l'enregistrament. Aquest software permet especificar el nombre de lectures que es fa per repetició, la resolució de les lectures, la utilització del mòdul a convenir i altres especificitats de cadascun d'aquests.



Mòdul d'esfera
amb rotació

Figura 12. Antaris II FT-NIR Analyzer. Fotografia: Max Michel

Les primeres lectures es fan amb la sonda de fibra òptica (figura 13) als tomàquets frescs, després d'haver estat rentats com s'ha comentat a l'apartat anterior. Per a cada fruit o mostra se'n fan 3 repeticions tenint en compte que per a cada repetició l'aparell realitza 16 lectures amb una resolució de 8cm^{-1} . El resultat obtingut de cada repetició és la mitjana de les 16 lectures, alhora les dades utilitzades posteriorment són la mitjana de les 3 repeticions.



Figura 13. Sonda de fibra òptica integrada a l'espectròmetre d'infraroig proper FT-NIR Analyzer. Fotografia: Max Michel

L'altre tipus de mostra que es registre són les farines de tomàquet. Després d'haver fet les mesures de la matèria seca, el tomàquet deshidratat restant a les safates d'alumini, ajuntant les dues repeticions, és molturat amb un molinet fins a convertir la matèria seca de tomàquet en una farina fina i homogènia (figura 14). Amb el mòdul de l'esfera amb rotació es va analitzar cadascuna de les farines fent 32 lectures cada vegada amb una resolució de 8cm^{-1} .



Figura 14. Farina de tomàquet. Fotografia: Max Michel

3.5. Anàlisi estadístic

Les dades obtingues s'han estudiat utilitzant el software R (R Core Team, 2017). S'ha tingut en compte els diferents genotips i tractaments com a factors per poder discernir si hi ha diferències significatives entre aquests. Primerament s'ha estudiat les dades més bàsiques: el rang de mesures, els màxims i mínims i les mitjanes i desviacions en les mesures dels diferents grups. Per a l'estudi de les diferències significatives entre grups s'ha utilitzat el procediment ANOVA en les dades de referència (anàlisis químics). Per als factors significatius ($p < 0,05$) s'ha fet un test HSD (de l'anglès "*Honestly Significant Difference*") de Tukey per a poder comparar de manera simultània diferents nivells dels dos factors a estudiar.

La informació obtinguda dels espectres dels fruits en fresc i d'aquest en farina van ser analitzats utilitzant el software PLS_Toolbox (MATLAB and Statistics Toolbox Release, 2012) per a dur a terme un anàlisi de regressió de mínims quadràtics parcials ("*Partial Least Square regression*"). Amb l'eina PLS_Toolbox s'ha pogut desenvolupar models de predicció utilitzant les dades espectrals enfront dels anàlisis de referència.

El primer pas necessari és el pretractament de les dades. En aquest treball s'ha provat tres tipus de pretractament: SNV ("*Standard Normal Variable*") i l'aplicació de la primera i la segona derivada, així com la combinació d'aquets pretractaments.. El desenvolupament de models predictius té tres passos clau: la calibració, la validació creuada ("*cross validation*") i la validació de la capacitat predictiva del model. Es comparen les dades espectrals amb les de referència. A partir de la matriu de dades aquesta és dividida en dues parts definint les dades segons si són per a fer la calibració i o per a fer la validació del model, l'algoritme de Kennard Stone divideix les mostres per a validar i per a calibrar basant-se en les diferències de probabilitats estocàstiques d'entre cada component amb els components circumdants. El procediment es fa per a la matriu de dades espectrals i de manera paral·lela a la matriu de dades de referència. Amb les matrius de calibració i utilitzant els diferents pretractaments esmentats es creen models de predicció. Aquets s'avaluen per una banda mitjançant en primer lloc, el *cross validation* en que la comprovació del model es fa a partir de mostres incloses en el model final i per altre banda amb la matriu de validació, on realment es comprova com es comporta el model creat. Per a cada part del procés es quantifica la validesa o encert del model resolt estudiant-ne el coeficient de determinació (R^2) i l'arrel quadràtica dels errors entre mitjanes (RMSE). A més per a cada model s'han utilitzat un nombre diferent de variables latents, variables que no han estat directament observades però que tenen inferència en el model estadístic. Tant per a l'elecció del tipus de pretractament dels

espectres, com el número de variables latents, s'ha tingut en compte que les R^2 fossin el més elevades possible i els RMSE el més baixos possible.

4. Resultats i discussió

4.1. Dades de referència

4.1.1. Estadístiques bàsiques

Abans de poder procedir amb l'estudi de diferències significatives entre diferents factors mitjançant el procediment de l'ANOVA s'ha estudiat la distribució de les dades obtingudes de les variables. Així doncs totes les variables estudiades mostren una distribució normal. A la figura 15 es poden observar els histogrames i la seva corba de distribució normal de 3 dels 5 caràcters estudiats.

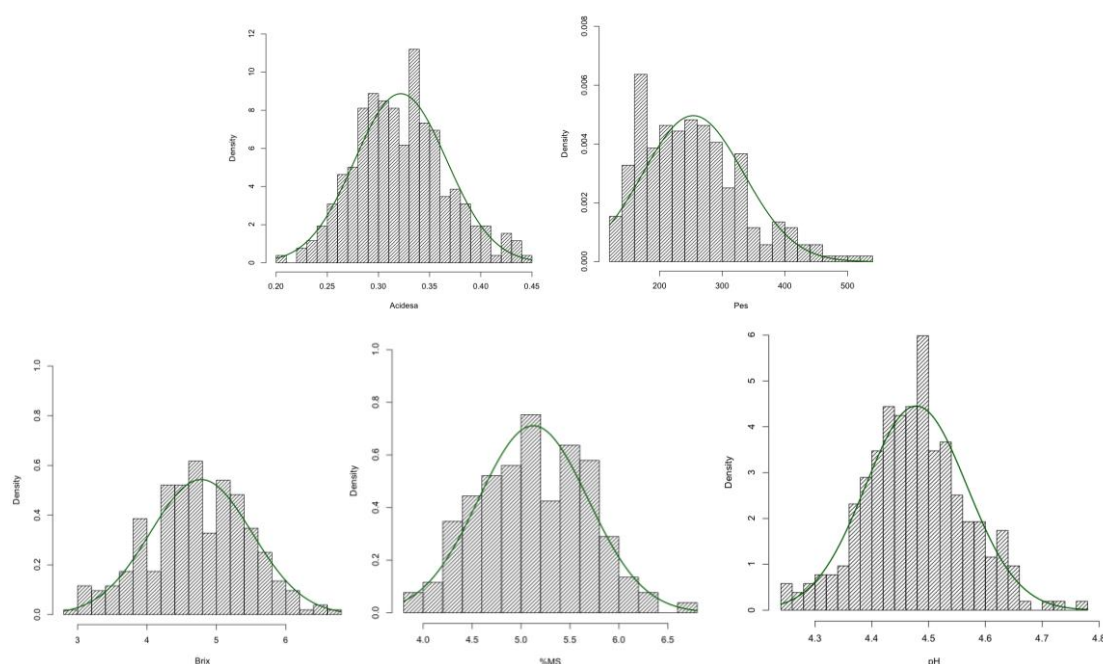


Figura 15. Histograma i corba de distribució normal dels °Brix, el percentatge en matèria seca i el pH.

La distribució de valors del percentatge de matèria seca, sense diferenciar el genotip o tractament, és bastant homogènia ja que el seu rang va des de 3,84% a 6,71%. Les mitjanes de les mesures d'aquest caràcter són diferenciables segons el seu genotip però sobretot destaca les diferències entre tractaments dins d'un mateix genotip, de manera que només en les varietats Guzman 8 i LC748 hi ha semblances de valors entre l'empeltat amb Beaufort i sense empelt i entre l'empeltat de Beaufort i Multifort respectivament (figura16).

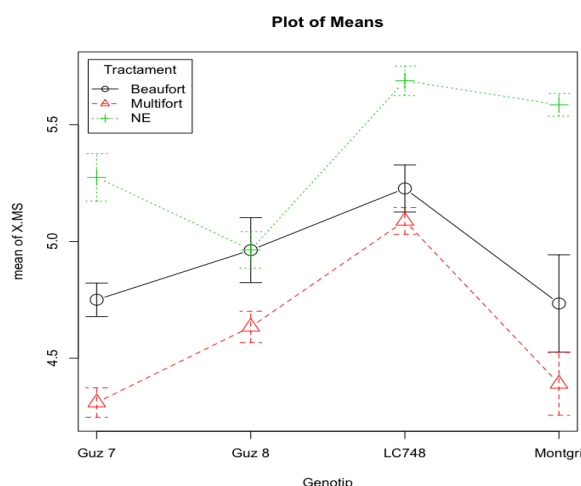


Figura 16. Distribució de les mitjanes i el rang des valors de percentatge en matèria seca en relació al genotip i tractament.

A la taula 3 es mostren els valors resum de les dades obtingudes en la fase experimental referents a la mesura del percentatge de matèria seca. En el rang de valors de les mitjanes no es pot apreciar una diferenciació clara i val tenir en compte que les dades de desviació dels resultats tenen valors molts petits significant un alt grau d'homogeneïtat de les dades. La diferència entre el valor màxim i el valor mínim en % de matèria seca és de 3 aproximadament, mostra senzilla altre cop de l'homogeneïtat de les mostres.

Taula 2. Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures de matèria seca en relació al genotip i tractament.

Genotip	Tractament	%MS,length	%MS(%),mean	%MS(%),sd	%MS(%),max	%MS(%),min
Guz 7	Beaufort	15	4,7501	0,2772	5,4587	4,3323
	Multifort	17	4,3096	0,2594	4,9732	3,8411
	NE	25	5,2748	0,5089	5,9837	4,1096
Guz 8	Beaufort	10	4,9627	0,4399	5,6715	4,1758
	Multifort	12	4,6340	0,2328	4,8229	4,0147
	NE	28	4,9641	0,4144	5,7384	4,2670
LC748	Beaufort	17	5,2271	0,4164	6,0440	4,4712
	Multifort	41	5,0881	0,3661	6,6895	4,5875
	NE	35	5,6886	0,3746	6,7091	5,0272
Montgrí	Beaufort	10	4,7344	0,6594	6,3947	4,2326
	Multifort	9	4,3897	0,4038	5,1568	3,9577
	NE	40	5,5857	0,3070	6,1172	4,7094
Nombre total de mostres		259				

A la figura 17 es pot observa els valors mitjos de pH. Aquests varien notablement entre genotips però poc entre tractaments. Concretament les mitjanes de pH de la varietat LC748 són més altes que la resta amb important predominança dels empeltats Beaufort i Multifort. En tots els genotips l'empeltat Multifort és determinant d'una mitja de pH major que la resta de tractaments sense importar la varietat en la que s'utilitzi aquest tractament.

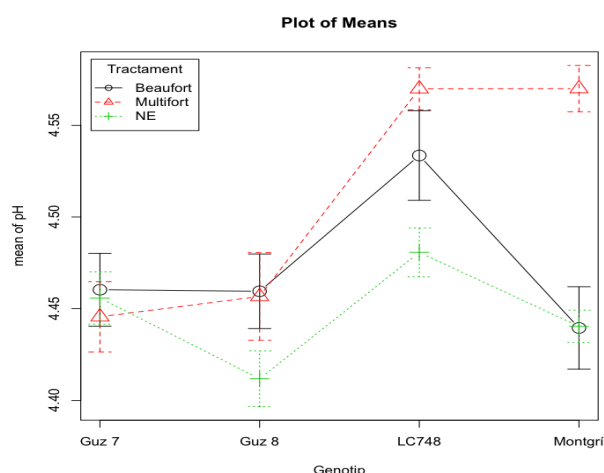


Figura 17. Distribució de les mitjanes i rang de valors de les mesures de pH en relació al genotip i tractament.

A la taula 3 es mostren els valors resum de les dades obtingudes en la fase experimental referents a la mesura de pH. Com era d'esperar, les diferències entre valors de pH són molt petites i cal tenir en compte l'escala de diferència entre valors. Com es pot apreciar a la figura 17 les diferències entre genotips és mínima, tenint en compte que la diferència de pH entre LC748, el genotip amb valors mitjans més elevats (sense comptar que concretament el genotip Montgrí amb empelt Beaufort té valors superiors) i Guzman 8, amb els valors mitjans més baixos, és d'entre 3,1 dècimes. Pel què fa als valors de les desviacions cal destacar diferències de valors entre genotips i tractaments així que el valor màxim i mínim disten de 7 centèsimes entre LC748 amb empelt Beaufort i Montgrí amb empelt Multifort respectivament.

Taula 3. Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures de pH en relació al genotip i tractament.

Genotip	Tractament	pH,length	pH,mean	pH,sd	pH,max	pH,min
Guz 7	Beaufort	15	4,4603	0,0770	4,5750	4,2850
	Multifort	17	4,4456	0,0791	4,5500	4,2550
	NE	25	4,4558	0,0715	4,6050	4,3150
Guz 8	Beaufort	10	4,4595	0,0642	4,5350	4,3700
	Multifort	12	4,4567	0,0827	4,5450	4,2400
	NE	28	4,4118	0,0804	4,5850	4,2450
LC748	Beaufort	17	4,5335	0,1006	4,7650	4,4000
	Multifort	41	4,5699	0,0734	4,7150	4,3950
	NE	35	4,4807	0,0788	4,6750	4,2750
Montgrí	Beaufort	10	4,4395	0,0711	4,5150	4,2700
	Multifort	9	4,5700	0,0378	4,6550	4,5350
	NE	40	4,4404	0,0553	4,5600	4,3250
Nombre total de mostres		259				

A la figura 18 es pot observar els valors mitjos de l'acidesa. S'aprecien subtils diferències depenent del tractament d'empeltat però la distinció per genotip és quasi nul·la. Només en el cas de la varietat LC748 les mostres tenen un rang de valor lleugerament més elevat que la resta (de només unes 2 mil·lèsimes). En qualsevol dels genotips les mitjanes d'acidesa valorable són sempre major en els fruits que no han rebut cap tractament d'empeltat. De manera contrària les mostres amb empelt Multifort, a excepció de la varietat Montgrí, tenen valors mitjans d'acidesa inferiors.

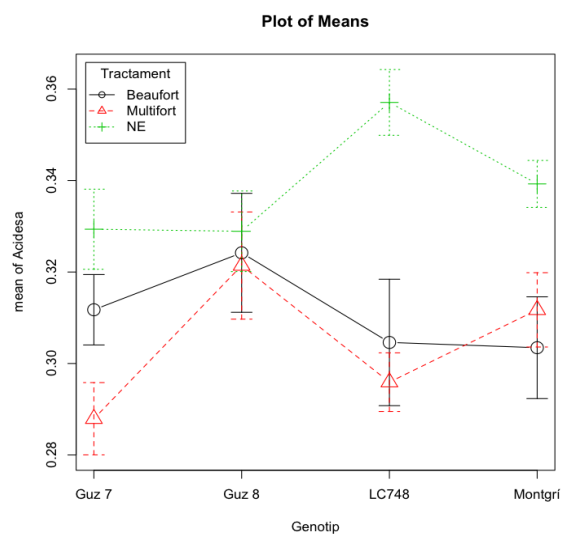


Figura 18. Distribució de les mitjanes i rang de valors de les mesures d'acidesa avalorable (TA) en relació al genotip i tractament.

A la taula 4 es mostren els valors resum de les dades obtingudes de la fase experimental referents a la mesura d'acidesa valorable. Els valors mitjans de cada genotip i tractament són bastant semblants i no disten més de 8 centèsimes entre el valor més alt i el més baix. Així els valors de desviació són mínims i els valors màxims i mínims de lectura són molt aproximats entre ells sense apreciar-se diferenciació entre genotips o tractaments.

Taula 4. Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures d'acidesa valorable (TA) en relació al genotip i tractament.

Genotip	Tractament	Nº Mostres	Acidesa(%),mean	Acidesa(%),sd	Acidesa(%),max	Acidesa(%),min
Guz 7	Beaufort	15	0,3117	0,0298	0,3614	0,2505
	Multifort	17	0,2879	0,0326	0,3631	0,2324
	NE	25	0,3294	0,0438	0,4482	0,2599
Guz 8	Beaufort	10	0,3242	0,0411	0,3816	0,2626
	Multifort	12	0,3214	0,0406	0,3869	0,2667
	NE	28	0,3289	0,0466	0,4336	0,2568
LC748	Beaufort	17	0,3046	0,0570	0,4382	0,2072
	Multifort	41	0,2959	0,0410	0,4055	0,2201
	NE	35	0,3571	0,0425	0,4379	0,2687
Montgri	Beaufort	10	0,3035	0,0352	0,3462	0,2448
	Multifort	9	0,3117	0,0244	0,3506	0,2849
	NE	40	0,3393	0,0325	0,4057	0,2796
Nombre total de mostres		259				

A la figura 19 es pot observar els valors mitjos del pes. Tant les diferències entre genotips com entre tractament són evidents a simple vista. Doncs clarament els valors de pes de la varietat LC748 són majors a la resta. A més, també clarament, en totes les varietats el tractament amb empelt Multifort són les mostres amb valors mitjos de pes majors. Respecte les varietats, en valor decreixent, l'ordre és: LC748, Guzman 8, Guzman 7 i Montgri. Pel que fa al tractament, també en valor decreixent, l'ordre és: Multifort, Beaufort i sense empelt. Segons la informació obtinguda de la figura 19 referent al valor de les mitjanes del pes, aquest

podria ser un caràcter determinant de la varietat i del tipus de tractament rebut en els fruits de tomàquet.

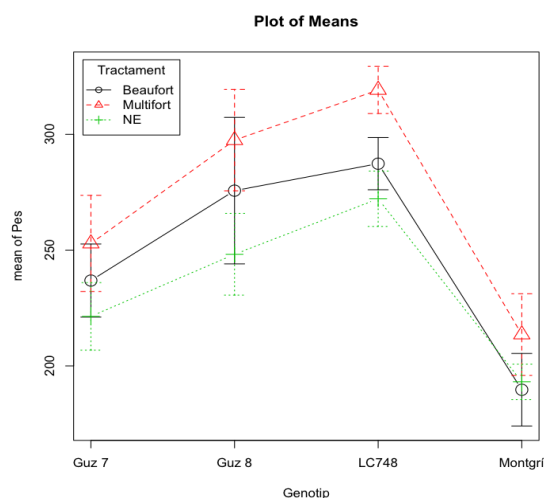


Figura 19. Distribució de les mitjanes i rang de valors de les mesures de pes en relació al genotip i tractament.

A la taula 5 es mostren els valors resum de les dades obtingudes de la fase experimental referents a la mesura del pes. Els valors de les mitjanes tenen diferències significatives entre ells tenint en compte que la distància màxim entre aquest és de 130g. Per tant els valors de desviació també són bastant elevats. Respecte els valors màxims i mínims tenir present que en la varietat LC748 amb tractament d'empelt Multifort el valor màxim de pes és de 526g i el mínim de 124g essent referent de l'ample rang de valors d'aquest caràcter també a tenir en compte que dins d'un mateix genotip la distància entre el valor màxim i mínim de pes és elevat.

Taula 5. Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures de pes en relació al genotip i tractament.

Genotip	Tractament	Nº Mostres	Pes(g),mean	Pes(g),sd	Pes(g),max	Pes(g),min
Guz 7	Beaufort	15	236,8667	61,1016	401	154
	Multifort	17	252,8824	85,6963	462	148
	NE	25	221,4000	72,9343	436	147
Guz 8	Beaufort	10	275,7000	100,0834	481	186
	Multifort	12	297,5000	75,9348	434	157
	NE	28	248,2143	93,3244	526	146
LC748	Beaufort	17	287,3529	46,6087	385	217
	Multifort	41	319,1951	65,1695	514	223
	NE	35	272,1714	70,7649	443	167
Montgrí	Beaufort	10	189,7000	49,5380	286	140
	Multifort	9	213,5556	52,9106	296	139
	NE	40	193,1250	48,4224	339	124
Nombre total de mostres		259				

A la figura 20 es pot observar els valors mitjos de °Brix. Les diferències entre genotips són observables, sent LC748 i Montgrí definits per mitjanes majors que la resta de genotips. En totes les varietats, els tomàquets que no han rebut cap tractament d'empeltat tenen valors mitjos de °Brix majors. Així, pel que fa al resum de mitjanes, els fruits sense tractament d'empelt tenen valors superiors als fruits tractats. Observant les mitjanes de les mostres

tractades amb empelt Beaufort i Multifort no sembla haver-hi un patró clar que defineixi cap diferència específica.

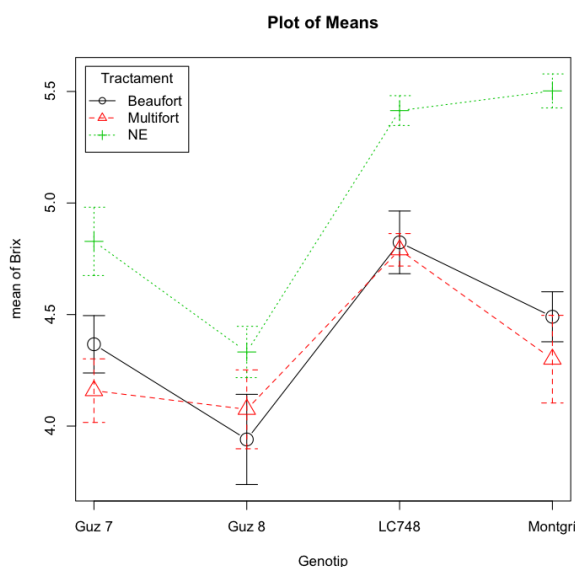


Figura 20. Distribució de les mitjanes i rang de valors de les mesures de sòlids solubles (°Brix) en relació al genotip i tractament.

A la taula 6 es mostren els valors resum de les dades obtingudes de la fase experimental referents a la mesura dels sòlids solubles (°Brix). De manera molt diferenciada els valors mitjans de les mostres sense empelt són majors. D'aquesta manera la menor diferència entre mitjanes és de 3 centèsimes i de 1,5° la major. És per això que els valors de desviació són relativament elevats, deguts a les diferències de les mitjanes respecte a les mitjanes de les mostres sense empelt. És interessant observar que mentre el valor màxim de °Brix mesurats és d'una mostra de la varietat LC748 sense empelt, com era d'esperar, la lectura amb valor mínim és de la varietat Guzman 8 sense empeltar.

Taula 6. Resum del nombre de mostres, mitjana, desviació, màxim i mínim de les mesures de sòlids solubles totals (°Brix) en relació al genotip i tractament.

Genotip	Tractament	Nº Mostres	Brix (°Bx),mean	Brix(°Bx),sd	Brix(°Bx),max	Brix(°Bx),min
Guz 7	Beaufort	15	4,3667	0,4981	4,9	3,3
	Multifort	17	4,1588	0,5874	5,5	2,8
	NE	25	4,8280	0,7662	6,2	3,2
Guz 8	Beaufort	10	3,9400	0,6398	5,5	3,2
	Multifort	12	4,0750	0,6122	5,1	3,1
	NE	28	4,3321	0,6092	5,5	3,1
LC748	Beaufort	17	4,8235	0,5794	5,9	3,9
	Multifort	41	4,7902	0,4657	5,6	3,7
	NE	35	5,4143	0,3964	6,7	4,7
Montgrí	Beaufort	10	4,4900	0,3542	5,2	3,9
	Multifort	9	4,3000	0,5895	5	3,3
	NE	40	5,5025	0,4822	6,5	4,6
Nombre total de mostres		259				

A la taula 7 es poden apreciar els valors de les mitjanes i desviacions entre genotips segons el procediment ANOVA. Segons aquest procediment és defineixen diferents grups de significació corresponents a grups de valors, lectures, diferenciats entre ells per a límits

establerts pel propi procés ANOVA. D'aquesta manera és possible estudiar el nivell de representació, significació, de les dades obtingudes i definir-les qualitativament.

Pel que fa als genotips, les diferenciacions són evidents en alguns casos com per exemple en el pes en que no hi ha cap repetició de grup entre els diferents genotips però, cal tenir en compte que és precisament per aquest paràmetre que s'observa una gran variància entre mostres d'un mateix genotip. Aquest alt valor de desviació entre resultats d'aquest paràmetre prové d'altres factors no estudiats en aquest treball. La definició de grups obtingut per als paràmetres d'acidesa valorable i pH no és estranya recapitulant la interpretació dels resultats de les mitjanes i desviacions, on les diferències entre els valors d'acidesa i pH són molt baixes.

Taula 7. Síntesi dels resultats obtinguts pel procediment ANOVA per a l'estudi entre genotips. Cada lletra ("a", "b" i "c") són els grups de significació, és a dir grups amb diferències delimitadores per al procediment, de manera que es pot establir ràpidament la capacitat de definició dels paràmetres respecte de les variables. La última fila presenta els valors de significació (valor p) dels factors considerats calculant la mitjana d'error quadràtic (*Mean Square Error*, la mitjana de les distàncies quadràtiques de cada mesura a la mitjana del grup).

	%MS		pH		Acidesa		Pes		Brix	
Guz 7	4,849±0,567	a	4,454±0,074	a	0,312±0,041	a	234,86±74,107	a	4,507±0,707	a
Guz 8	4,885±0,403	a	4,432±0,08	b	0,326±0,043	a	265,54±91,521	ab	4,192±0,626	a
LC748	5,339±0,466	b	4,53±0,09	b	0,321±0,053	a	295,677±67,413	bc	5,031±0,548	b
Montgrí	5,259±0,626	b	4,46±0,072	b	0,329±0,035	a	195,661±49,029	c	5,147±0,705	b
valor p	0,15		0,005		0,002		4846,135		0,297	

Les diferències entre diferents tractaments són poc significatives. Per al percentatge en matèria seca i pH cada tractament i l'absència d'aquest pertany cadascun a un grup estadístic concret. L'acidesa valorable, el pes i els °Brix tenen la mateixa distribució de grups on el tractament amb empelt Beaufort es diferencia de el tractament amb empelt Multifort i en absència d'empelt.

Taula 8. Síntesi dels resultats obtinguts pel procediment ANOVA per a l'estudi entre tractaments d'empeltat. Cada lletra ("a", "b" i "c") són els grups de significació, és a dir grups amb diferències delimitadores per al procediment, de manera que es pot establir ràpidament la capacitat de definició dels paràmetres respecte de les variables. La última fila presenta els valors de significació (valor p) dels factors considerats calculant la mitjana d'error quadràtic (*Mean Square Error*, la mitjana de les distàncies quadràtiques de cada mesura a la mitjana del grup).

	%MS		pH		Acidesa		Pes		Brix	
Beaufort	4,944±0,483	a	4,48±0,089	a	0,31±0,043	a	251,769±72,541	a	4,458±0,606	a
Multifort	4,772±0,475	b	4,526±0,092	b	0,3±0,039	b	289,595±78,84	b	4,49±0,61	b
NE	5,417±0,48	c	4,448±0,075	c	0,34±0,042	b	232,313±76,93	b	5,091±0,725	b
valor p	0,15		0,005		0,002		4846,135		0,297	

4.2. Anàlisi de l'espectre NIR

4.2.1. Espectres sense tractar

A partir de la recopilació de les dades obtingudes amb l'espectròmetre FT-NIR s'ha obtingut el conjunt d'espectre dels tomàquets sencer registrats amb pistola (figura 21) i de les mostres seques (en forma de farina homogènia) (figura 22). Sense més manipulació no es possible apreciar diferències significatives entre els diferents tipus de mostra ja que les iteracions, o lectures, per a cada interval de longitud d'ona són molt petites. Per això és necessari fer el pretractament de les dades i posteriors estudis estadístics. A partir de les gràfiques tampoc és possible establir patrons clars de diferenciació però si que es pot apreciar que, si més no,

els espectres no es solapen pel que el prosseguir amb l'estudi estadístic està completament avalat.

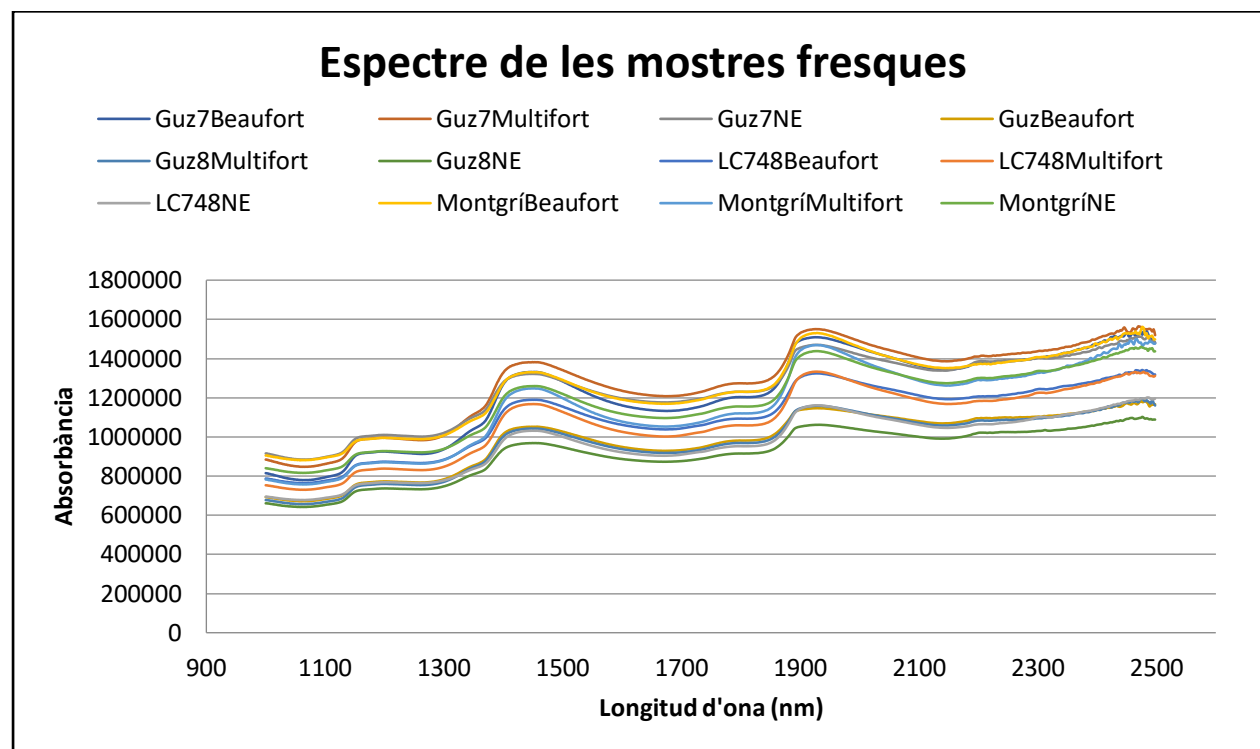


Figura 21. Espectres sense tractar de les mostres molturades diferenciades per genotip i tractament.

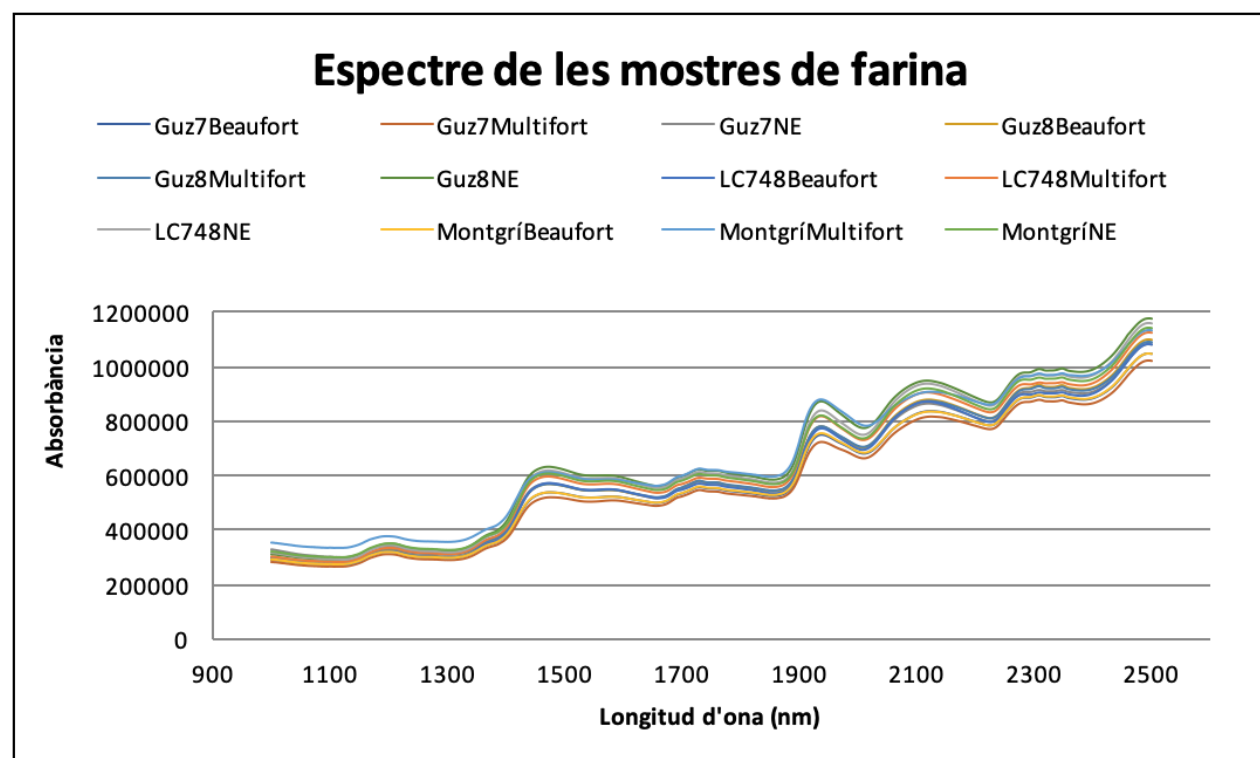


Figura 22. Espectres sense tractar de les mostres fresques diferenciades per genotip i tractament.

4.2.2. Regressió de PLS

En el procediment de regressió PLS (*Partial Least Square*) s'ha pogut determinar la relació entre les diferents variables tenint en compte les dues maneres de registre de les mostres. Així s'ha estudiat la calibració de les matrius emprades, s'ha fet un *cross-validation* i a posteriorment s'ha estudiat la capacitat de predicció dels models a partir de la matriu de validació. A la figura 23 es pot observar la distribució de les mostres estudiades. En el registre en fresc la distribució no mostra una resposta adequada a la modelització de regressió lineal i de manera que no es possible definir un patró per a la caracterització espectral de les mostres segons el seu genotip. Contràriament les en el registre amb farina de tomàquet les mostres responen a la modelització de regressió lineal i disposen un patró lineal respecte dels eixos establerts de manera que és possible discernir entre diferents genotips a partir de l'espectre electromagnètic.

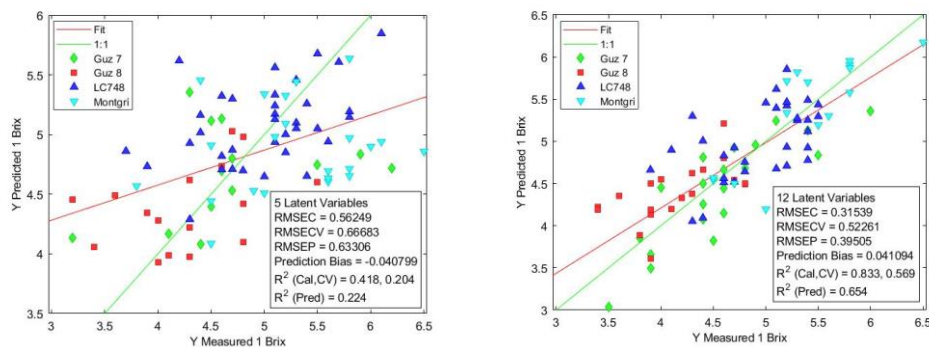


Figura 23. Representació gràfica de la distribució de mostres segons el procediment de regressió per PLS per la variable °Brix segons genotip i tractament per al registre en fresc o en farina.

A la taula 9 es pot observar els resultats obtinguts d'aquest procediment de manera que és fàcil veure que els percentatges de validesa dels models de predicció són massa baixos per a categoritzar-los com a vàlids. Els models de predicció a partir del procediment de regressió de mínims quadràtics parcials (PLS) no són aptes per a cap dels paràmetres estudiats. L'únic model que té un coeficient de determinació òptim és per a la quantitat de °Brix en el registre amb farines. Aquest és valuós i és mostra de la bona praxis en l'estudi però el fet que sigui en el registre en farina és important. En el sector l'estudi d'aquest paràmetre es fa sempre en fresc i és relativament senzill de fer-ne lectures si es disposa d'un refractòmetre.

En altres estudis en que s'han desenvolupat models de predicció òptims a partir de l'espectre electromagnètic de tomàquets frescs les mostres utilitzades són fruits amb molt poques diferències entre ells. De manera que el model final obtingut es refereix a un sol genotip de tomàquet i aquest modificat genèticament per tenir en majoria fruits de característiques molt semblants. Els models predictius no han estat bons perquè hi ha massa variància entre els fruits estudiats, especialment en el registre en fresc utilitzant la sonda de fibra òptica. Per tant la síntesis de models de predicció per a tomàquets és possible però és necessari disposar d'un conjunt de mostres més homogeni, és a dir, entre fruits que siguin més semblants física i químicament. Aquest és el cas dels treballs nombrats a l'apartat 1.3.3. on tots els estudis han utilitzat el mostres de tomàquet de la mateixa espècie i amb morfologies semblants, menys el cas en que s'estudia tenint en compte diferents estats de maduresa (Huang et al., 2018). A més el nombre de fruits utilitzats és menor en tots ells al nombre utilitzat en aquest treball.

Taula 9. Síntesi dels resultats obtinguts pel procediment de regressió de mínims quadràtics parcials (PLS regression).

				Calibració	Cross-validation			Predicció	
Paràmetre	Registre	Pretractament	LVs	R ²	RMSEC	R ²	RMSECV	R ²	RMSEP
°Brix	Fresc	SNV+1derivada	5	0,41	0,56	0,2	0,67	0,22	0,63
%MS	Fresc	1derivada	5	0,38	0,45	0,17	0,53	0,24	0,45
Acidesa	Fresc	SNV	3	0,18	0,04	0,09	0,04	0,02	0,04
°Brix	Farina	SNV+1derivada	12	0,83	0,32	0,57	0,52	0,65	0,4
Acidesa	Farina	SNV	11	0,44	0,03	0,26	0,04	0,28	0,04

5. Conclusions

Els paràmetres de qualitat química del tomàquet estudiats en aquest treball són el pes, el percentatge en matèria seca, el pH, l'acidesa valorable (TA) i els sòlids solubles totals (°Brix). Els anàlisis de referència demostren que la distribució de resultats de tots els paràmetres segueix una distribució de dades normal. El procediment ANOVA per als paràmetres mostra diferències significatives entre diferents grups segons diferenciació per genotip o tractament.

En aquest treball, segons el seguiment del disseny experimental i el rang de mostres utilitzats, no es pot determinar positivament la capacitat de l'espectroscòpia NIR per a mesurar paràmetres de qualitat química del tomàquet. És possible relacionar els espectres electromagnètics obtinguts amb el NIR i els anàlisis de referència però els nivells d'acceptació dels models emprats són molt baixos. Les diferències entre genotips i tractaments són observables en els anàlisis de referència però no ho són de manera concloent en els models desenvolupats amb els espectres enregistrats, on els valors d'acceptació són molt baixos.

Els models de predicció obtinguts no són aptes i tenen valors d'acceptació molt baixos. D'altre banda, fent palesa de la capacitat de l'espectroscòpia NIR per al desenvolupament de models predictius de paràmetres de qualitat del tomàquet, el model desenvolupat per als °Brix en farina té valors d'acceptació òptims i defineix una capacitat de modelització lineal dels espectres en relació a les dades de referència òptima.

6. Bibliografia

- Abney, W.; Festing, E.R. (1881) On the influence of the Atomic Grouping in the Molecules of Organic Bodies on Their Absorption in the infra-Red Region of the Spectrum. *Philosophical Transactions of the Royal Society of London*. doi:10.1098/rspl.1880.0056
- Baker, M. J., Trevisan, J., Bassan, P., Bhargava, R., Butler, H. J., Dorling, K. M., ... Martin, F. L. (2014). Using Fourier transform IR spectroscopy to analyze biological materials, 9(8), 1771–1791. <http://doi.org/10.1038/nprot.2014.110>
- Belmes, K. B., & Stauffer, C. E. (2018). Using a Michelson Interferometer to Measure the Index of Refraction of Air Using a Michelson Interferometer to Measure the Index of Refraction of, 3.
- Blanco, M., & Villarroya, I. (2002). NIR spectroscopy: A rapid-response analytical tool. *TrAC - Trends in Analytical Chemistry*, 21(4), 240–250. [http://doi.org/10.1016/S0165-9936\(02\)00404-1](http://doi.org/10.1016/S0165-9936(02)00404-1)
- Clément, A., Dorais, M., & Vernon, M. (2008). Multivariate approach to the measurement of tomato maturity and gustatory attributes and their rapid assessment by vis-NIR spectroscopy. *Journal of Agricultural and Food Chemistry*, 56(5), 1538–1544. <http://doi.org/10.1021/jf072182n>
- Deák, K. J., Szigedi, T., Palotás, G., Daood, H. G., & Helyes, L. (2015). Determination of °Brix, lycopene, β -carotene and total carotenoid content of processing tomatoes using near infrared spectroscopy. *Acta Horticulturae*, 1081, 253–258. <http://doi.org/10.17660/ActaHortic.2015.1081.32>
- Do electromagnetic waves occupy varying amounts of space, or do they simply vary in magnitude?. (2015-03-19. Joshua Lin. *Physics Questions*. <https://physics.stackexchange.com/questions/171144/do-electromagnetic-waves-occupy-varying-amounts-of-space-or-do-they-simply-vary>
- Donald A. Burns, Emil W. Ciurczak., 2007. *Handbook of Near-Infrared Analysis*. Third edition. London. CRC Press, 2007. ISBN: 978-0-8493-7393-0.
- Ecarnot, M., Baogonekczyk, P., Tessarotto, L., & Chervin, C. (2013). Rapid phenotyping of the tomato fruit model, Micro-Tom, with a portable VIS-NIR spectrometer. *Plant Physiology and Biochemistry*, 70, 159–163. <http://doi.org/10.1016/j.plaphy.2013.05.019>
- Engelhard, Y. N., Gazer, B., & Paran, E. (2006). Natural antioxidants from tomato extract reduce blood pressure in patients with grade-1 hypertension: A double-blind, placebo-controlled pilot study. *American Heart Journal*, 151(1), 1–6. <http://doi.org/10.1016/j.ahj.2005.05.008>
- Frusciante, L., Carli, P., Ercolano, M. R., Pernice, R., Di Matteo, A., Fogliano, V., & Pellegrini, N. (2007). Antioxidant nutritional quality of tomato. *Molecular Nutrition and Food Research*, 51(5), 609–617. <http://doi.org/10.1002/mnfr.200600158>
- Huang, Y., Lu, R., Xu, Y., & Chen, K. (2018). Prediction of tomato firmness using spatially-resolved spectroscopy. *Postharvest Biology and Technology*, 140(January), 18–26. <http://doi.org/10.1016/j.postharvbio.2018.02.008>
- Huck, C. W. (2017). Theoretical and technical advancements of near-infrared spectroscopy and its operational impact in industry. *NIR News*, 28(8), 17–21. <http://doi.org/10.1177/0960336017738174>

- Jordan, Edward C., K. G. B. (1968). Electromagnetic Waves and Radiation Systems. New Jersey: Prentice-Hall Electrical engineering series.
- Kaye, W. (1954). Near-infrared spectroscopy: I. Spectral identification and analytical applications. *Spectrochimica Acta*, 6(4), 257--E2. [http://doi.org/10.1016/0371-1951\(54\)80011-7](http://doi.org/10.1016/0371-1951(54)80011-7)
- Khuwijtjaru, P. (2018). Near infrared spectroscopy research performance in food science and technology. *NIR News*, The Author(s) 2018. <http://doi.org/10.1177/0960336018763197>.
- Kónya, J., & Nagy, N. M. (2012). 5 - Interaction of Radiation with Matter. *Nuclear and Radiochemistry*, 83--127. <http://doi.org/https://doi.org/10.1016/B978-0-12-391430-9.00005-6>
- Lee, H., Kim, M. S., Jeong, D., Delwiche, S. R., Chao, K., & Cho, B. K. (2014). Detection of cracks on tomatoes using a hyperspectral near-infrared reflectance imaging system. *Sensors (Switzerland)*, 14(10), 18837--18850. <http://doi.org/10.3390/s141018837>
- MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.
- Murray, I. 2004. Scattered information: philosophy and practice of near infrared spectroscopy. *Proceedings of the 11th International Conference on Near Infrared spectroscopy*. NIR Publications 2.
- Norris, K. H. (1964). Simple spectroradiometer for 0.4 to 1.2-micron region. *Transactions of the ASAE*, 7, 240--242.
- Norris, K. H., & Butler, W. L. (1961). Techniques for obtaining absorption spectra on intact biological samples. *IRE Transactions on Bio-Medical Electronics*, 8(C), 153--157. <http://doi.org/10.1109/TBMEL.1961.4322890>
- Nuez, F., 1995. *El cultivo del tomate*. Madrid: Ediciones Mundi-Prensa. ISBN 8471145499.
- Brian C. Smith., 2011. *Fundamentals of Fourier transform infrared spectroscopy*. Second edition. London. CRC Press, 2011. ISBN: 978-1-4200-6930-3.
- Plans, M. (2014). Utilització de l'espectroscòpia d'infraroig com a eina per a la selecció i millora de la mongeta seca (*Phaseolus vulgaris* L.). Retrieved from <http://www.tdx.cat/handle/10803/283234>
- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Reich, G. (2005). Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews*, 57(8), 1109--1143. <http://doi.org/10.1016/j.addr.2005.01.020>
- Saad, A., Jha, S. N., Jaiswal, P., Srivastava, N., & Helyes, L. (2016). Non-destructive quality monitoring of stored tomatoes using VIS-NIR spectroscopy. *Engineering in Agriculture, Environment and Food*, 9(2), 158--164. <http://doi.org/10.1016/j.eaef.2015.10.004>
- Skoog, D.A., Holler, F.J. and Nieman, T.A. 2001. *Principios de analisis instrumental*. McGraw-Hill. Madrid. Spain.

Workman Jr., J. (1995). Vibrational spectroscopy. *Vibrational Spectroscopy*, 8(1&2), 291–302.
<http://doi.org/10.1016/j.vibspec.2011.02.005>